Supplemental Materials on

Single-Image Depth Estimation Based on Fourier Domain Analysis

Jae-Han Lee, Minhyeok Heo, Kyung-Rae Kim, and Chang-Su Kim
Korea University
{jaehanlee, mhheo, krkim}@mcl.korea.ac.kr, changsukim@korea.ac.kr

S-1. Depth Map Refinement

Optionally, we refine local details of a depth map $\hat{\mathbf{D}}$ using superpixels. An estimated depth map is often blurry around depth discontinuities. To reduce such artifacts, we use the color information in an image, since depth discontinuities occur at color edges in general. To extract color edges, we use the LSC superpixel method [S1].

We generate $\hat{\mathbf{D}}_n^{\mathrm{s}}$ and $\hat{\mathbf{D}}_n^{\sigma}$ using the superpixel segmentation result of an input image, where n is the number of superpixels. $\hat{\mathbf{D}}_n^{\mathrm{s}}$ has an identical depth within each superpixel, which is the mean of depth values in the corresponding area of $\hat{\mathbf{D}}$. Similarly, $\hat{\mathbf{D}}_n^{\sigma}$ records the standard deviations of depths within superpixels. We use the standard deviations to guide the refinement and penalize depths significantly deviated from mean values.

The number n of superpixels should be selected appropriately. If n is too large, there are too many superpixels of small sizes and the mean estimation becomes unreliable. On the contrary, if n is too small, each superpixel covers a big area, and the assumption that all its pixels have the same depth becomes invalid. Considering this tradeoff, we use four numbers $n_1 = 40$, $n_2 = 80$, $n_3 = 160$ and $n_4 = 320$ to refine $\hat{\mathbf{D}}$ into the final depth map $\tilde{\mathbf{D}}$.

To refine the depth of each pixel $\mathbf{x}$, we first calculate $\hat{\mathbf{D}}^{\mathrm{s}}$ by averaging all $\hat{\mathbf{D}}_n^{\mathrm{s}}$. Also, we obtain the standard score map $\hat{\mathbf{D}}^{\mathrm{z}}$ as follows.

$$\hat{\mathbf{D}}^{\mathrm{z}}(\mathbf{x}) = \frac{1}{4}\sum_{k=1}^{4} \frac{\hat{\mathbf{D}}(\mathbf{x}) - \hat{\mathbf{D}}_{n_k}^{\mathrm{s}}(\mathbf{x})}{\hat{\mathbf{D}}_{n_k}^{\sigma}(\mathbf{x})}. \tag{1}$$

If $\hat{\mathbf{D}}^{\mathrm{z}}$ exceeds an upper bound or a lower bound, we penalize the depth value. We set these bounds to 0.5 and -0.5, and define the depths corresponding to these bound values as $\hat{\mathbf{D}}^{\mathrm{over}}$ and $\hat{\mathbf{D}}^{\mathrm{under}}$, respectively. Then, we penalize the depths, which exceed bounds, by

$$\tilde{\mathbf{D}}(\mathbf{x}) \tag{2}$$
$$= \begin{cases} \frac{\hat{\mathbf{D}}(\mathbf{x}) + \hat{\mathbf{D}}^{\mathrm{under}}(\mathbf{x})}{2} & \text{if } \hat{\mathbf{D}}(\mathbf{x}) < \hat{\mathbf{D}}^{\mathrm{under}}(\mathbf{x}), \\ \frac{\hat{\mathbf{D}}(\mathbf{x}) + \hat{\mathbf{D}}^{\mathrm{over}}(\mathbf{x})}{2} & \text{if } \hat{\mathbf{D}}(\mathbf{x}) > \hat{\mathbf{D}}^{\mathrm{over}}(\mathbf{x}), \\ \hat{\mathbf{D}}(\mathbf{x}) & \text{otherwise.} \end{cases}$$

This refinement technique improves depth estimation results quantitatively and qualitatively.

S-2. More Ablation Study

The proposed algorithm has three key components:

- DEN: depth estimation network, modified from ResNet-152 [11].
- DBE: depth-balanced Euclidean loss for estimating shallow depths more reliably.
- FDC: Fourier domain combination of multiple depth map candidates.

Also, we denote the depth map refinement process in Section S-1 as 'REF.' We apply the components sequentially to our network 'DEN' and also to three popular networks AlexNet [16], VGG19 [S2], and ResNet-52 [11].

Table S-1 shows that each component contributes to the performance improvement. The only exception is that DBE does not improve AlexNet. This is because, as compared with the other networks, AlexNet yields relative large errors for distant objects, as well as for near objects. Thus, balancing the loss for shallow depths is not effective in this case. The results in Table S-1 indicate that each component is not only effective for the proposed DEN, but also for the other networks. Therefore, these components can be used for other single-image depth estimation techniques as well.

In Table S-2, we compare the proposed algorithm including 'REF' with the conventional algorithms [1, 3, 5, 6, 19-21, 28, 38, 43]. We see that the proposed algorithm still achieves competitive depth estimation performances, even when DEN is replaced by VGG19 or ResNet-50, which have fewer layers and parameters. Also, note that [3] and [19] are the state-of-art algorithms, which are based on VGG19 and ResNet-52, respectively. Compared to these algorithms, 'VGG19+DBE+FDC(18)+REF' and 'ResNet-52+DBE+FDC(18)+REF' provide better performances.

Figure S-1 compares depth maps qualitatively. Again, when the components DBE, FDC, and REF are incorpo-

Table S-1. The three components DBE, FDC, and REF of the proposed algorithm improve the depth estimation performances of AlexNet, VGG19, and ResNet-50, as well as those of DEN.

	The lower, the better				The higher, the better		
	RMSE (lin)	RMSE (log)	Abs Rel	Sqr Rel	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
AlexNet	0.836	0.296	0.244	**0.238**	60.4%	87.6%	**96.4%**
AlexNet+DBE	0.870	0.299	**0.243**	0.244	60.3%	87.1%	96.3%
AlexNet+DBE+FDC(18)	0.826	**0.295**	0.247	0.243	61.5%	87.7%	96.2%
AlexNet+DBE+FDC(18)+REF	**0.825**	**0.295**	0.247	0.242	**61.6%**	**87.8%**	96.2%
VGG19	**0.616**	0.213	0.163	0.120	76.6%	95.2%	98.9%
VGG19+DBE	0.619	0.211	0.158	0.116	76.9%	95.3%	99.0%
VGG19+DBE+FDC(18)	0.617	**0.209**	**0.157**	**0.113**	77.0%	**95.4%**	99.0%
VGG19+DBE+FDC(18)+REF	0.617	**0.209**	**0.157**	**0.113**	**77.2%**	**95.4%**	**99.1%**
ResNet-50	**0.591**	0.203	0.151	0.108	79.4%	95.5%	99.0%
ResNet-50+DBE	0.603	0.203	0.147	0.106	**79.6%**	95.6%	99.0%
ResNet-50+DBE+FDC(18)	0.597	0.201	0.145	0.102	79.3%	95.7%	**99.1%**
ResNet-50+DBE+FDC(18)+REF	0.596	**0.200**	**0.144**	**0.101**	79.4%	**95.8%**	**99.1%**
DEN	0.586	0.199	0.145	0.104	80.3%	96.1%	99.0%
DEN+DBE	0.585	0.196	0.142	0.102	81.3%	96.1%	98.9%
DEN+DBE+FDC(18)	0.572	0.193	**0.139**	0.096	81.5%	**96.3%**	**99.1%**
DEN+DBE+FDC(18)+REF	**0.570**	**0.192**	**0.139**	**0.095**	**81.7%**	**96.3%**	**99.1%**

Table S-2. Performance comparison of the proposed algorithm, the three network-substituted versions of the proposed algorithm, and the conventional algorithms. The best results are boldfaced, and the second best ones are underlined. Here, 'DEN+DBE+FDC(18)+REF' is the proposed algorithm.

	The lower, the better				The higher, the better		
	RMSE (lin)	RMSE (log)	Abs Rel	Sqr Rel	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Zoran *et al*. [43]	1.220	0.430	0.410	0.570	-	-	-
AlexNet+DBE+FDC(18)+REF	0.825	0.295	0.247	0.242	61.6%	87.8%	96.2%
Li *et al*. [20]	0.821	-	0.232	-	62.1%	88.6%	96.8%
Liu *et al*. [21]	0.824	-	0.230	-	61.4%	88.3%	97.1%
Baig *et al*. [1]	0.802	-	0.241	-	61.0%	-	-
Eigen *et al*. [6]	0.877	0.283	0.214	0.204	61.4%	88.8%	97.2%
Wang *et al*. [38]	0.745	0.262	0.220	0.210	60.5%	89.0%	97.0%
Roy *et al*. [28]	0.744	-	0.187	-	-	-	-
Eigen and Fergus [5]	0.641	0.214	0.158	0.121	76.9%	95.0%	98.8%
Chakrabarti *et al*. [3]	0.620	0.205	0.149	0.118	80.6%	95.8%	98.7%
VGG19+DBE+FDC(18)+REF	0.617	0.209	0.157	0.113	77.2%	95.4%	**99.1%**
Laina *et al*. [19]	0.597	0.204	0.140	0.106	81.1%	95.3%	98.8%
ResNet-50+DBE+FDC(18)+REF	0.596	0.200	0.144	0.101	79.4%	95.8%	**99.1%**
DEN+DBE+FDC(18)+REF	**0.570**	**0.192**	**0.139**	**0.095**	**81.7%**	**96.3%**	**99.1%**

rated, AlexNet, VGG19, and ResNet-52 yield higher quality depth maps. Compared with the results using the networks only, the results using the three components estimate the depths more accurately and also yield less blurring artifacts.

Finally, Figure S-2 compares the results of the proposed algorithm 'DEN+DBE+FDC(18)+REF' with those of the network-substituted versions. We observe that the proposed algorithm yields the best results by employing a more effective network. However, even with VGG19 or ResNet-52, the proposed algorithm yields decent depth maps.

References

[S1] J. Chen, Z. Li, and B. Huang. Linear spectral clustering superpixel. *IEEE Trans. Image Process.*, 26(7):3317–3330, 2017.

[S2] K. Simonyan and A. Zisserman. *Very deep convolutional networks for large-scale image recognition*. [Online]. Available: https://arxiv.org/abs/1409.1556.

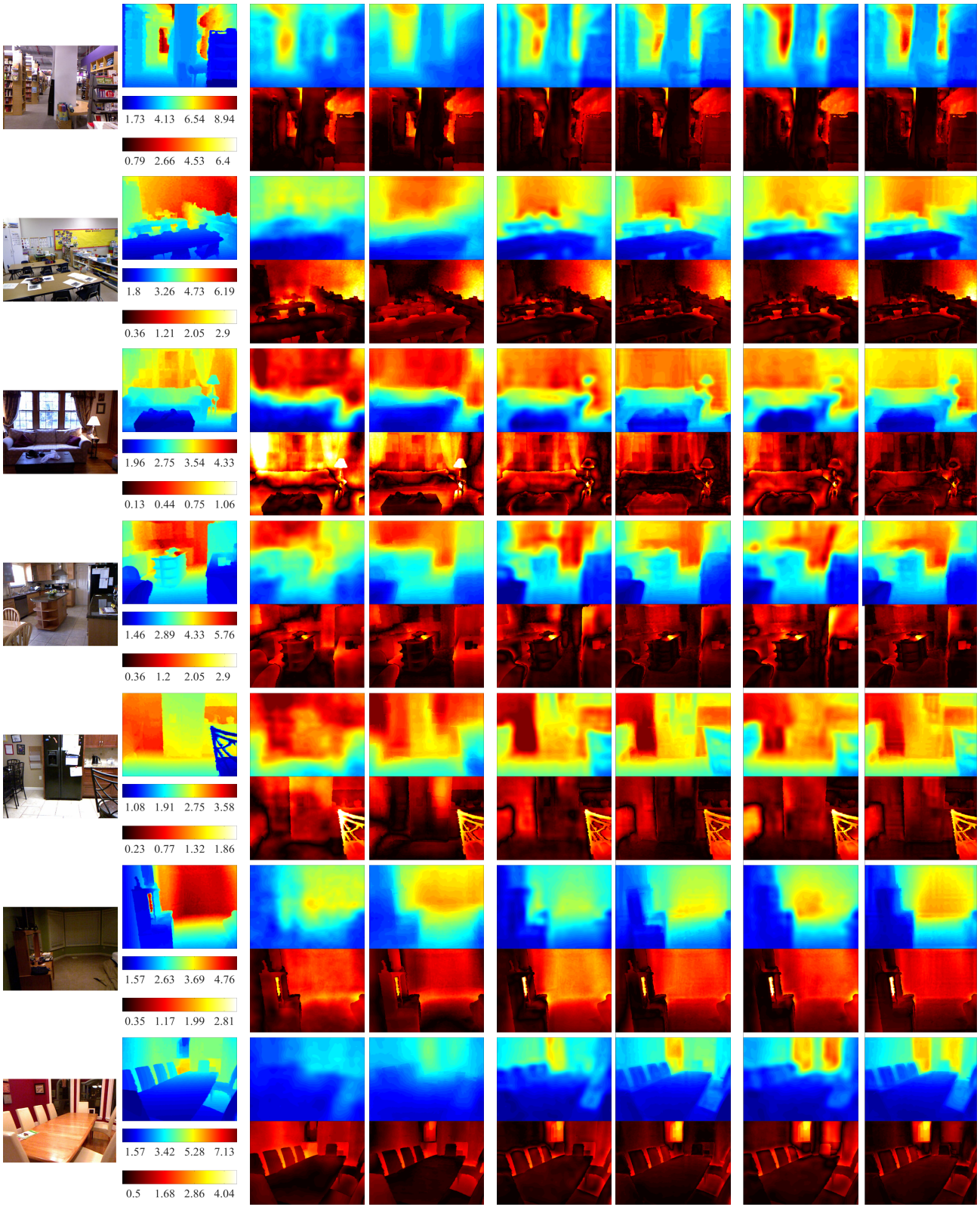

Figure S-1. Comparison of estimated depth maps. Upper images show depth maps, and lower images are the corresponding error maps. (a) Input images, (b) ground-truth depth maps and color coding schemes, (c) AlexNet, (d) AlexNet+DBE+FDC(18)+REF, (e) VGG19, (f) VGG19+DBE+FDC(18)+REF, (g) ResNet-52, and (h) ResNet-52+DBE+FDC(18)+REF.

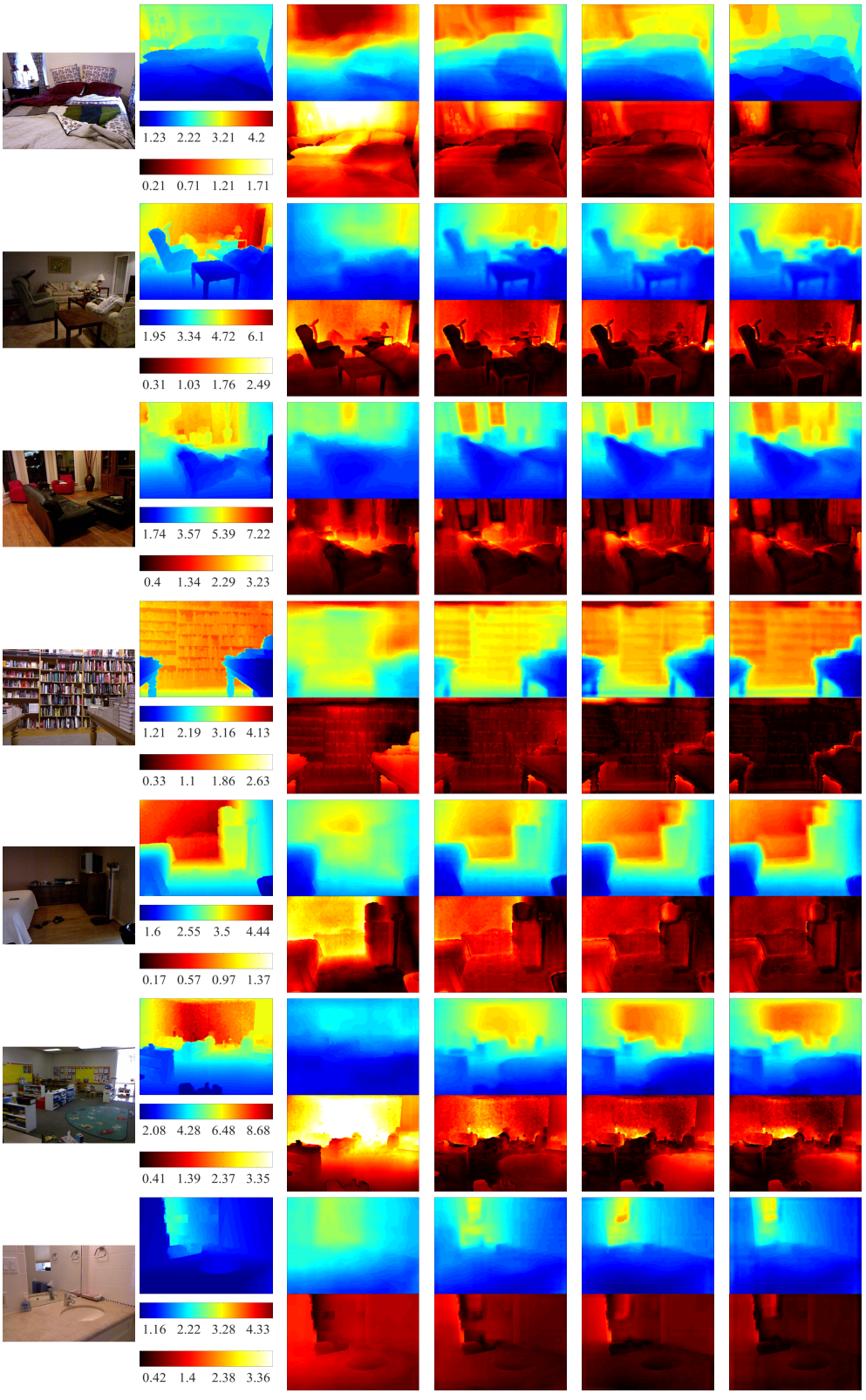

Figure S-2. Comparison of estimated depth maps: (a) input images, (b) ground-truth depth maps and color coding schemes, (c) AlexNet+DBE+FDC(18)+REF, (d) VGG19+DBE+FDC(18)+REF, (e) ResNet-52+DBE+FDC(18)+REF, and (f) DEN+DBE+FDC(18)+REF.